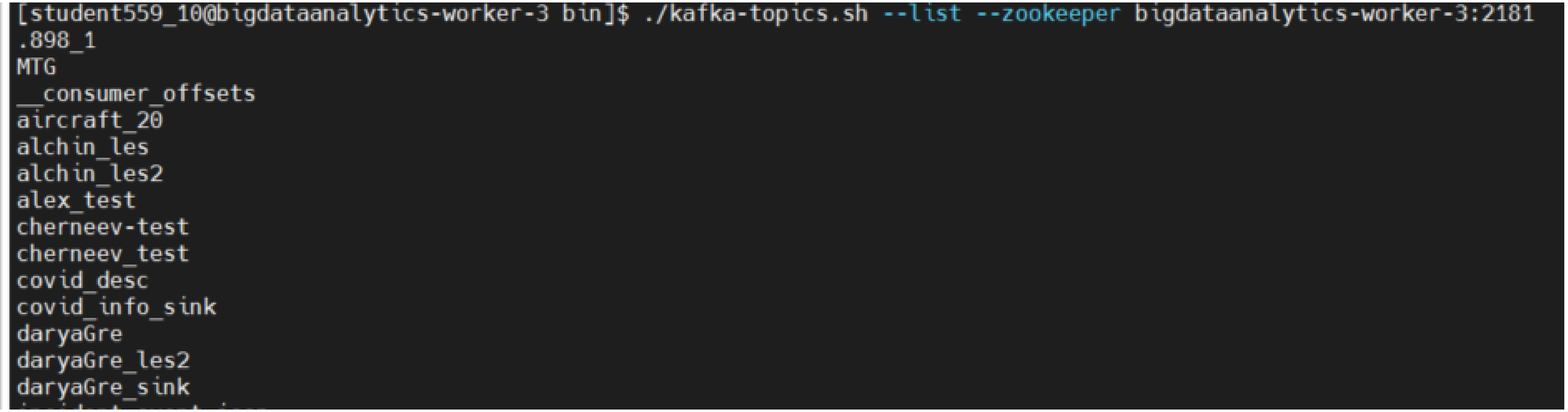
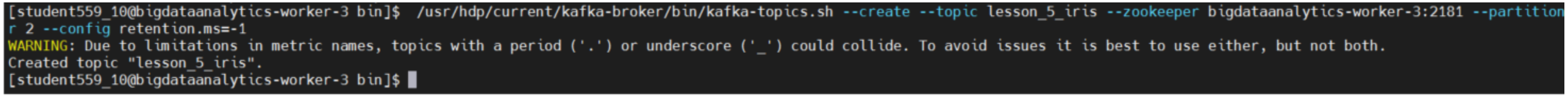
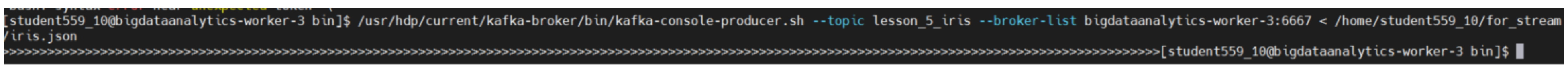


```
[student559_10@bigdataanalytics-worker-3 bin]$ ./kafka-topics.sh --list --zookeeper bigdataanalytics-worker-3:2181
.898_1
MTG
__consumer_offsets
aircraft_20
alchin_les
alchin_les2
alex_test
cherneev-test
cherneev_test
covid_desc
covid_info_sink
daryaGre
daryaGre_les2
daryaGre_sink
```



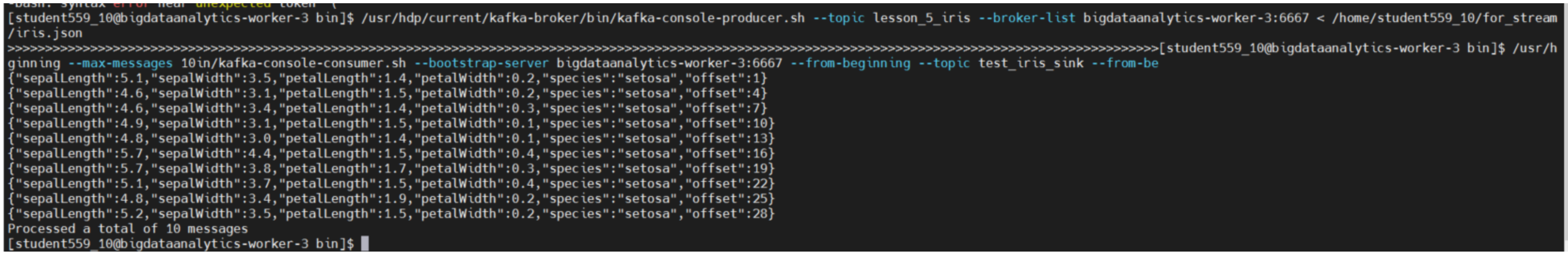
```
[student559_10@bigdataanalytics-worker-3 bin]$  /usr/hdp/current/kafka-broker/bin/kafka-topics.sh --create --topic lesson_5_iris --zookeeper bigdataanalytics-worker-3:2181 --partition
r 2 --config retention.ms=-1
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issues it is best to use either, but not both.
Created topic "lesson_5_iris".
[student559_10@bigdataanalytics-worker-3 bin]$ █
```

ort MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net

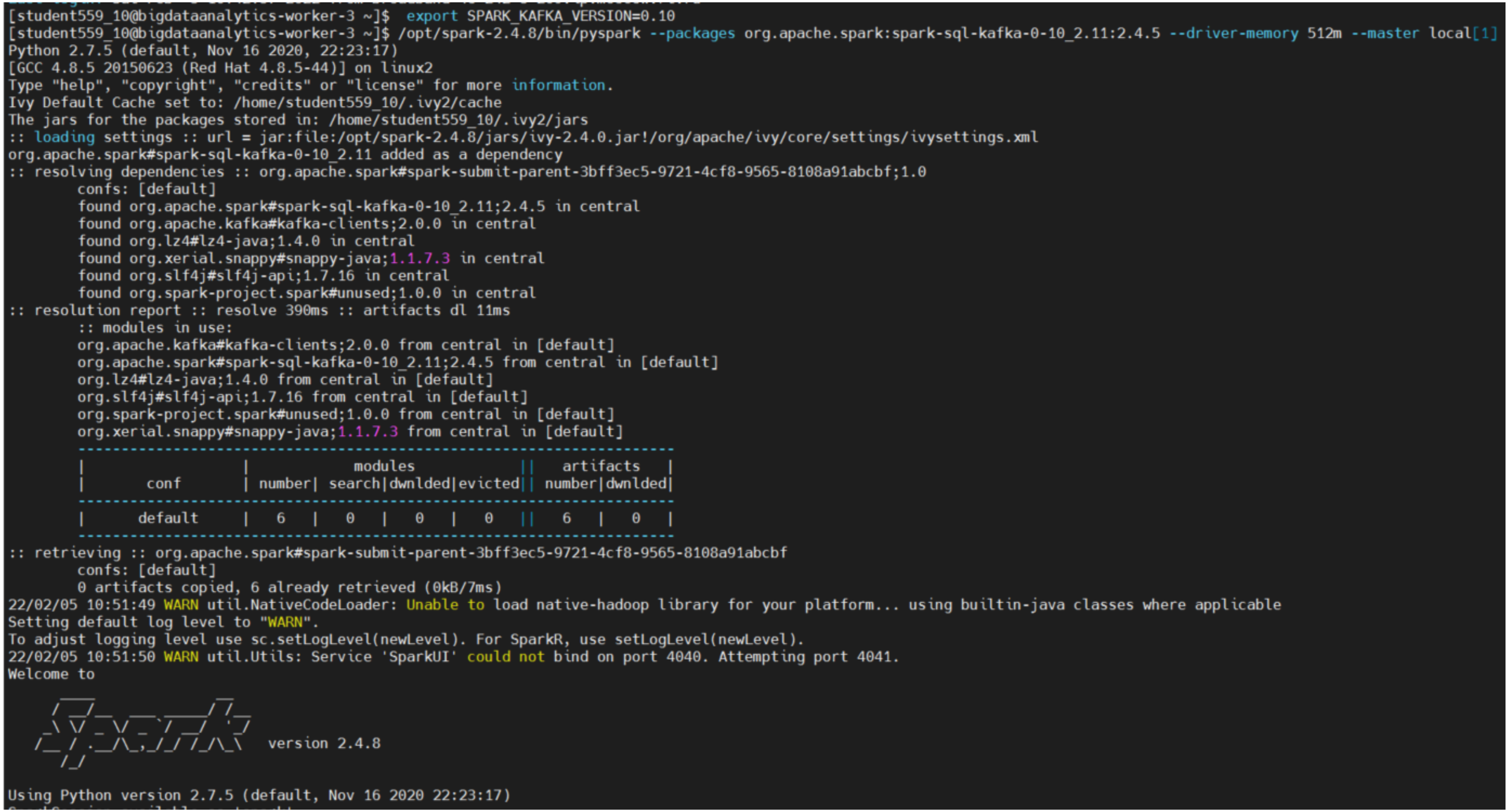


```
-bash: syntax error near unexpected token `('
[student559_10@bigdataanalytics-worker-3 bin]$ /usr/hdp/current/kafka-broker/bin/kafka-console-producer.sh --topic lesson_5_iris --broker-list bigdataanalytics-worker-3:6667 < /home/student559_10/for_stream
/iris.json
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>[student559_10@bigdataanalytics-worker-3 bin]$ █
```

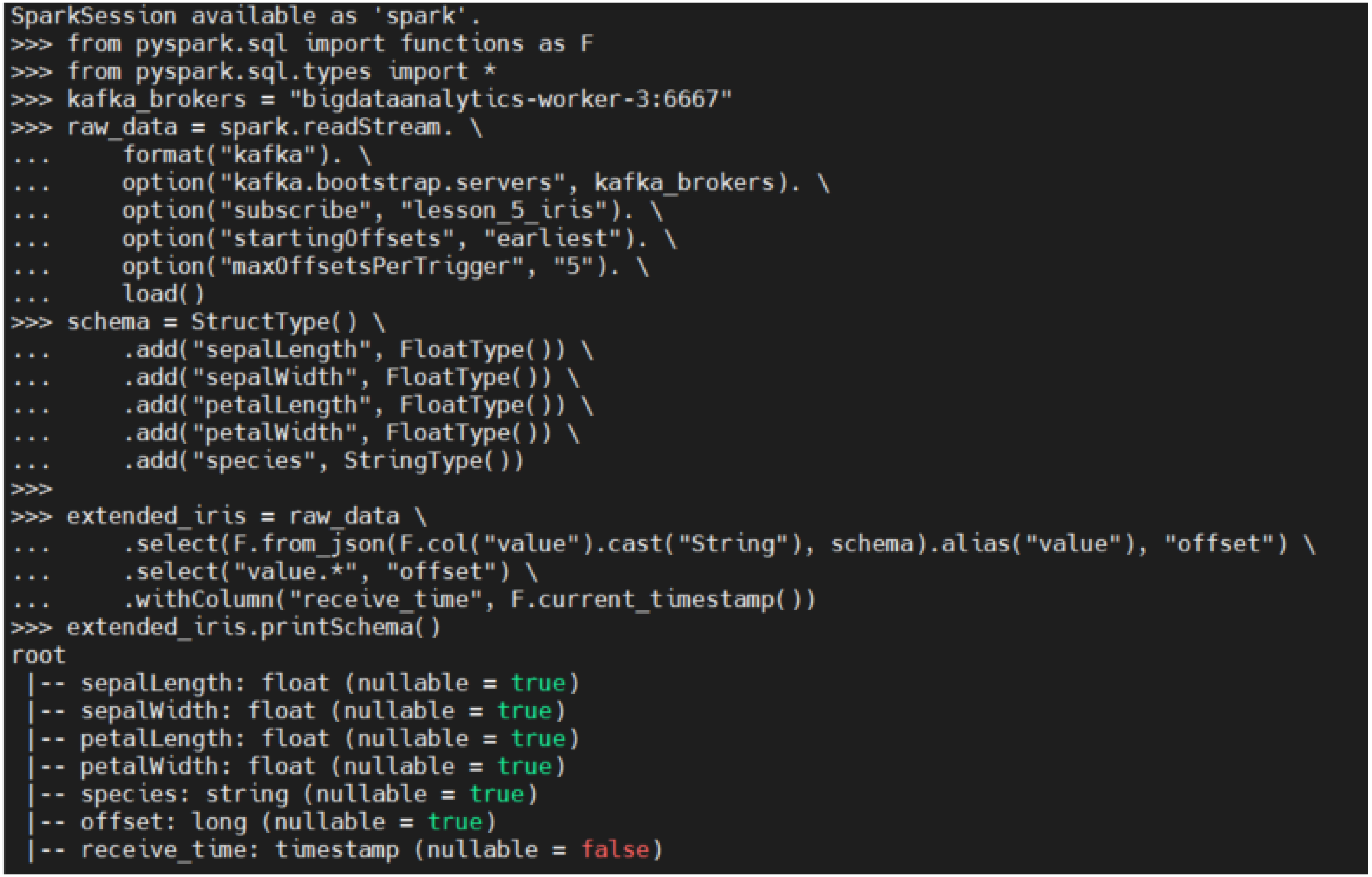
rt MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net



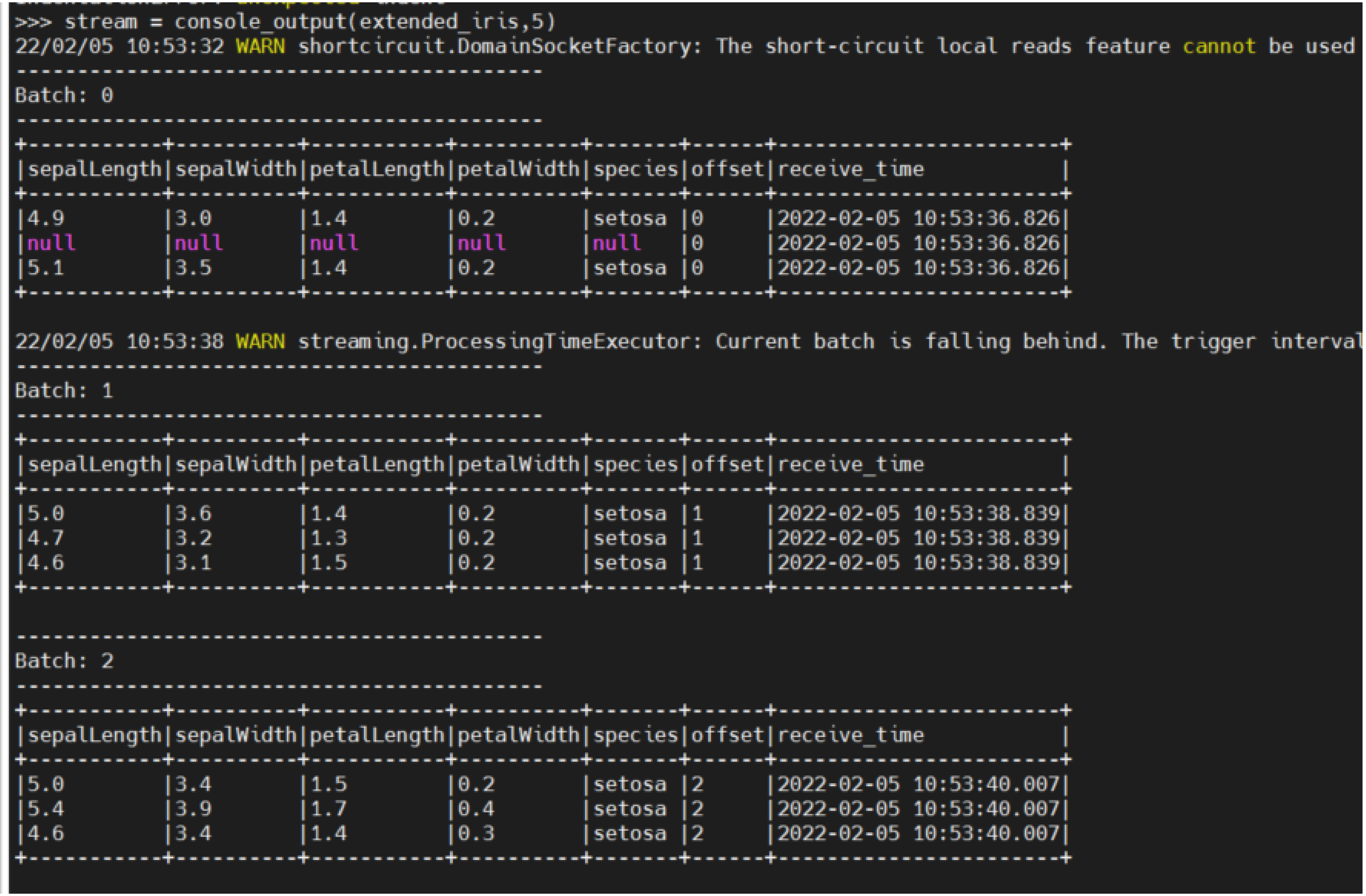
```
-bash: syntax error near unexpected token `('
[student559_10@bigdataanalytics-worker-3 bin]$ /usr/hdp/current/kafka-broker/bin/kafka-console-producer.sh --topic lesson_5_iris --broker-list bigdataanalytics-worker-3:6667 < /home/student559_10/for_stream
/iris.json
>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>[student559_10@bigdataanalytics-worker-3 bin]$ /usr/h
ginning --max-messages 10in/kafka-console-consumer.sh --bootstrap-server bigdataanalytics-worker-3:6667 --from-beginning --topic test_iris_sink --from-be
{"sepalLength":5.1,"sepalWidth":3.5,"petalLength":1.4,"petalWidth":0.2,"species":"setosa","offset":1}
{"sepalLength":4.6,"sepalWidth":3.1,"petalLength":1.5,"petalWidth":0.2,"species":"setosa","offset":4}
{"sepalLength":4.6,"sepalWidth":3.4,"petalLength":1.4,"petalWidth":0.3,"species":"setosa","offset":7}
{"sepalLength":4.9,"sepalWidth":3.1,"petalLength":1.5,"petalWidth":0.1,"species":"setosa","offset":10}
{"sepalLength":4.8,"sepalWidth":3.0,"petalLength":1.4,"petalWidth":0.1,"species":"setosa","offset":13}
{"sepalLength":5.7,"sepalWidth":4.4,"petalLength":1.5,"petalWidth":0.4,"species":"setosa","offset":16}
{"sepalLength":5.7,"sepalWidth":3.8,"petalLength":1.7,"petalWidth":0.3,"species":"setosa","offset":19}
{"sepalLength":5.1,"sepalWidth":3.7,"petalLength":1.5,"petalWidth":0.4,"species":"setosa","offset":22}
{"sepalLength":4.8,"sepalWidth":3.4,"petalLength":1.9,"petalWidth":0.2,"species":"setosa","offset":25}
{"sepalLength":5.2,"sepalWidth":3.5,"petalLength":1.5,"petalWidth":0.2,"species":"setosa","offset":28}
Processed a total of 10 messages
[student559_10@bigdataanalytics-worker-3 bin]$ █
```



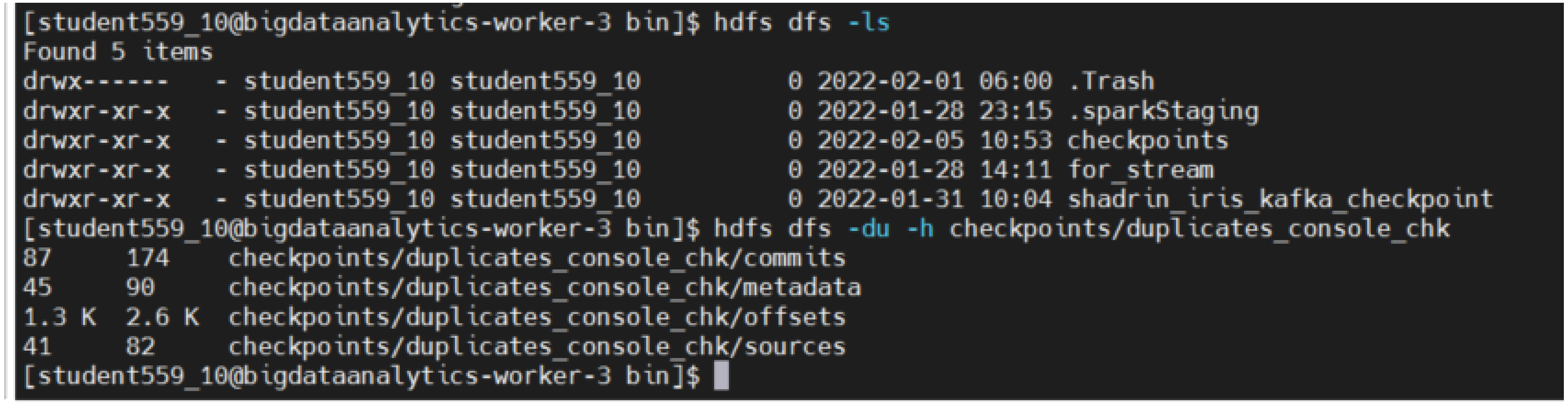
```
[student559_10@bigdataanalytics-worker-3 ~]$  export SPARK_KAFKA_VERSION=0.10
[student559_10@bigdataanalytics-worker-3 ~]$ /opt/spark-2.4.8/bin/pyspark --packages org.apache.spark:spark-sql-kafka-0-10_2.11:2.4.5 --driver-memory 512m --master local[1]
Python 2.7.5 (default, Nov 16 2020, 22:23:17)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-44)] on linux2
Type "help", "copyright", "credits" or "license" for more information.
Ivy Default Cache set to: /home/student559_10/.ivy2/cache
The jars for the packages stored in: /home/student559_10/.ivy2/jars
:: loading settings :: url = jar:file:/opt/spark-2.4.8/jars/ivy-2.4.0.jar!/org/apache/ivy/core/settings/ivysettings.xml
org.apache.spark#spark-sql-kafka-0-10_2.11 added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-3bff3ec5-9721-4cf8-9565-8108a91abcbf;1.0
        confs: [default]
        found org.apache.spark#spark-sql-kafka-0-10_2.11;2.4.5 in central
        found org.apache.kafka#kafka-clients;2.0.0 in central
        found org.lz4#lz4-java;1.4.0 in central
        found org.xerial.snappy#snappy-java;1.1.7.3 in central
        found org.slf4j#slf4j-api;1.7.16 in central
        found org.spark-project.spark#unused;1.0.0 in central
:: resolution report :: resolve 390ms :: artifacts dl 11ms
        :: modules in use:
        org.apache.kafka#kafka-clients;2.0.0 from central in [default]
        org.apache.spark#spark-sql-kafka-0-10_2.11;2.4.5 from central in [default]
        org.lz4#lz4-java;1.4.0 from central in [default]
        org.slf4j#slf4j-api;1.7.16 from central in [default]
        org.spark-project.spark#unused;1.0.0 from central in [default]
        org.xerial.snappy#snappy-java;1.1.7.3 from central in [default]
        ---------------------------------------------------------------------
        |                  |            modules            ||   artifacts   |
        |       conf       | number| search|dwnlded|evicted|| number|dwnlded|
        ---------------------------------------------------------------------
        |      default     |   6   |   0   |   0   |   0   ||   6   |   0   |
        ---------------------------------------------------------------------
:: retrieving :: org.apache.spark#spark-submit-parent-3bff3ec5-9721-4cf8-9565-8108a91abcbf
        confs: [default]
        0 artifacts copied, 6 already retrieved (0kB/7ms)
22/02/05 10:51:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
22/02/05 10:51:50 WARN util.Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 2.4.8
      /_/

Using Python version 2.7.5 (default, Nov 16 2020 22:23:17)
```



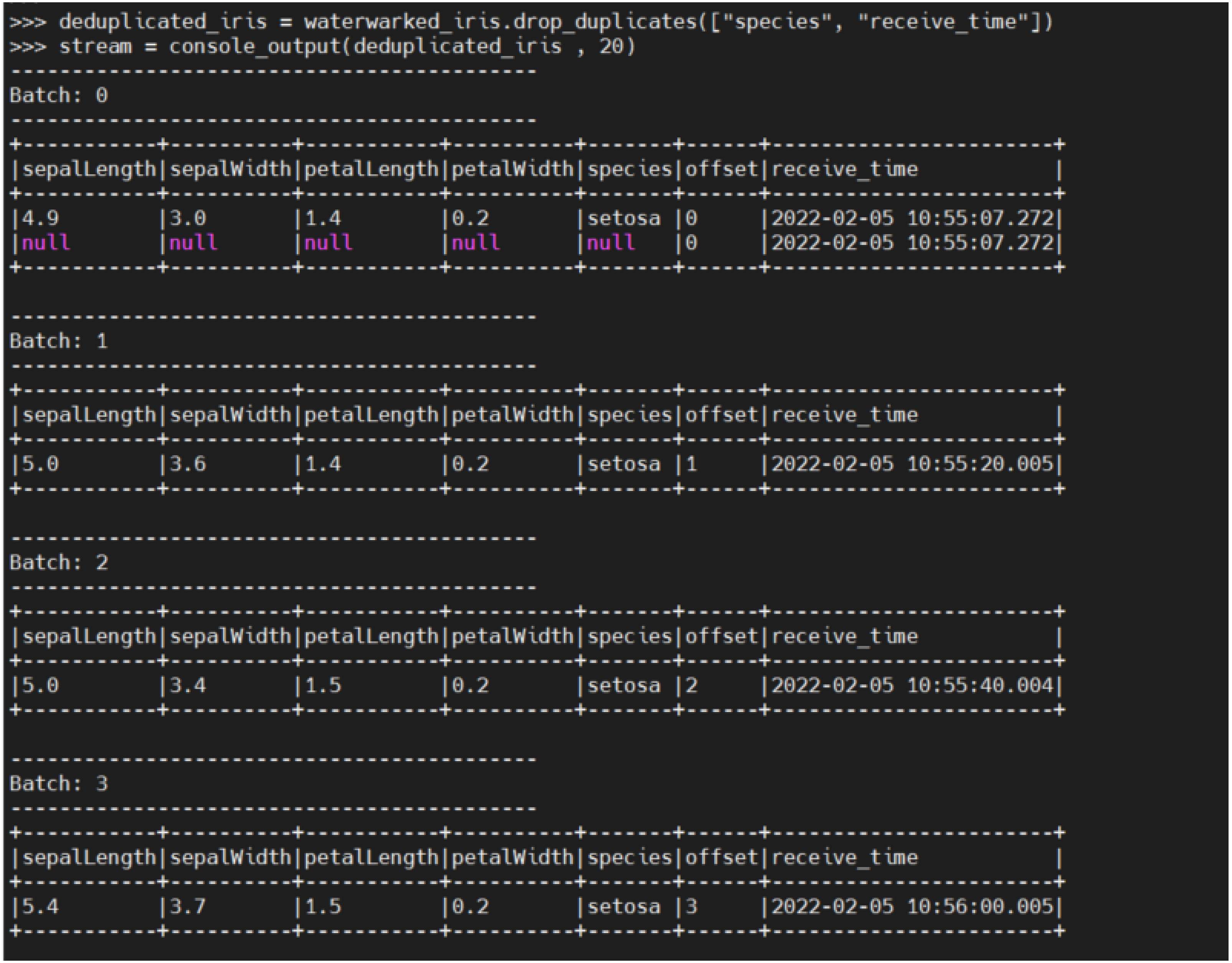
```
SparkSession available as 'spark'.
>>> from pyspark.sql import functions as F
>>> from pyspark.sql.types import *
>>> kafka_brokers = "bigdataanalytics-worker-3:6667"
>>> raw_data = spark.readStream. \
...     format("kafka"). \
...     option("kafka.bootstrap.servers", kafka_brokers). \
...     option("subscribe", "lesson_5_iris"). \
...     option("startingOffsets", "earliest"). \
...     option("maxOffsetsPerTrigger", "5"). \
...     load()
>>> schema = StructType() \
...     .add("sepalLength", FloatType()) \
...     .add("sepalWidth", FloatType()) \
...     .add("petalLength", FloatType()) \
...     .add("petalWidth", FloatType()) \
...     .add("species", StringType())
>>>
>>> extended_iris = raw_data \
...     .select(F.from_json(F.col("value").cast("String"), schema).alias("value"), "offset") \
...     .select("value.*", "offset") \
...     .withColumn("receive_time", F.current_timestamp())
>>> extended_iris.printSchema()
root
 |-- sepalLength: float (nullable = true)
 |-- sepalWidth: float (nullable = true)
 |-- petalLength: float (nullable = true)
 |-- petalWidth: float (nullable = true)
 |-- species: string (nullable = true)
 |-- offset: long (nullable = true)
 |-- receive_time: timestamp (nullable = false)
```



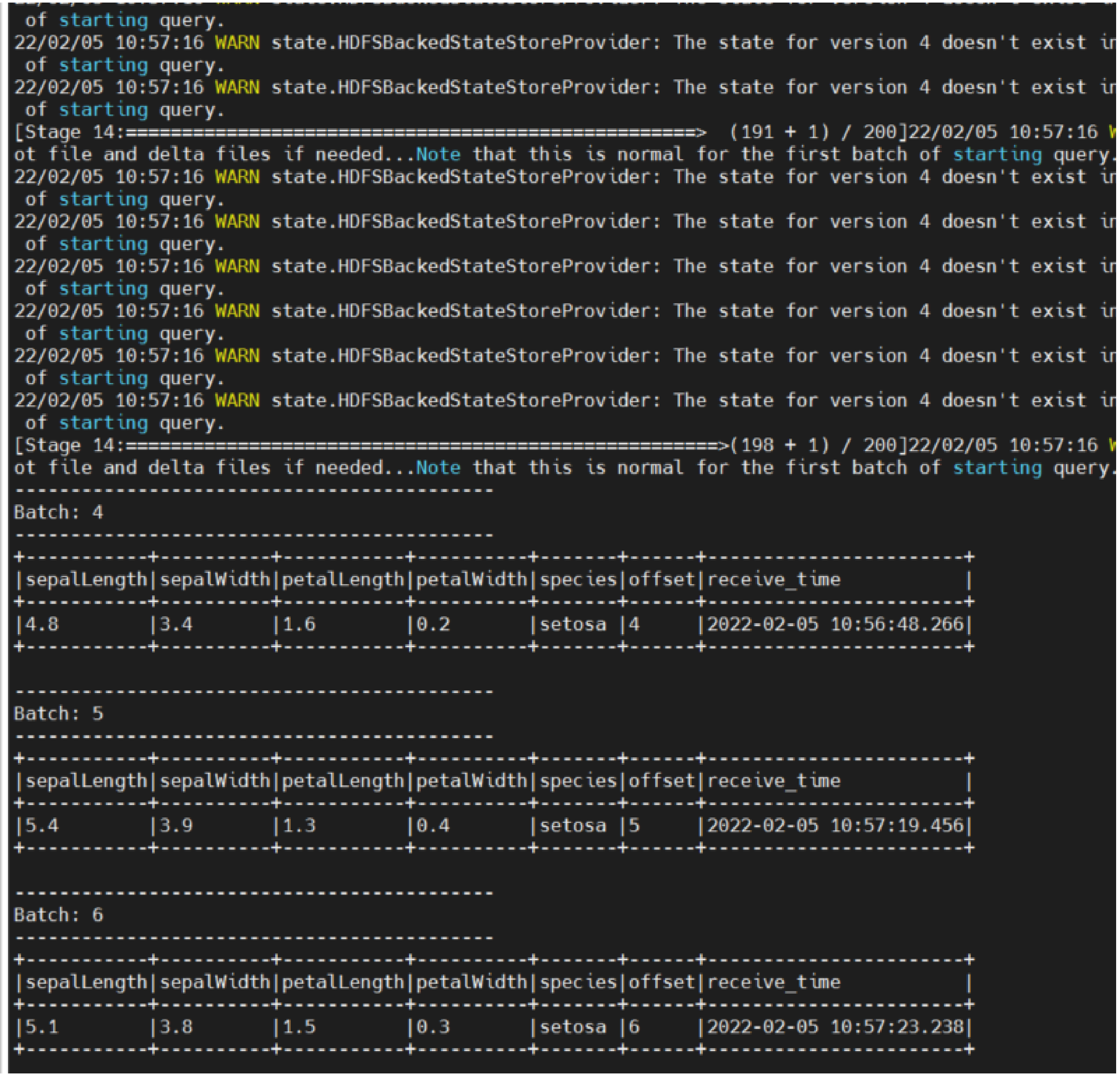
```
>>> stream = console_output(extended_iris,5)
22/02/05 10:53:32 WARN shortcircuit.DomainSocketFactory: The short-circuit local reads feature cannot be used
-------------------------------------------
Batch: 0
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|4.9        |3.0       |1.4        |0.2       |setosa |0     |2022-02-05 10:53:36.826|
|null       |null      |null       |null      |null   |0     |2022-02-05 10:53:36.826|
|5.1        |3.5       |1.4        |0.2       |setosa |0     |2022-02-05 10:53:36.826|
+-----------+----------+-----------+----------+-------+------+-----------------------+

22/02/05 10:53:38 WARN streaming.ProcessingTimeExecutor: Current batch is falling behind. The trigger interval
-------------------------------------------
Batch: 1
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|5.0        |3.6       |1.4        |0.2       |setosa |1     |2022-02-05 10:53:38.839|
|4.7        |3.2       |1.3        |0.2       |setosa |1     |2022-02-05 10:53:38.839|
|4.6        |3.1       |1.5        |0.2       |setosa |1     |2022-02-05 10:53:38.839|
+-----------+----------+-----------+----------+-------+------+-----------------------+


-------------------------------------------
Batch: 2
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|5.0        |3.4       |1.5        |0.2       |setosa |2     |2022-02-05 10:53:40.007|
|5.4        |3.9       |1.7        |0.4       |setosa |2     |2022-02-05 10:53:40.007|
|4.6        |3.4       |1.4        |0.3       |setosa |2     |2022-02-05 10:53:40.007|
+-----------+----------+-----------+----------+-------+------+-----------------------+
```



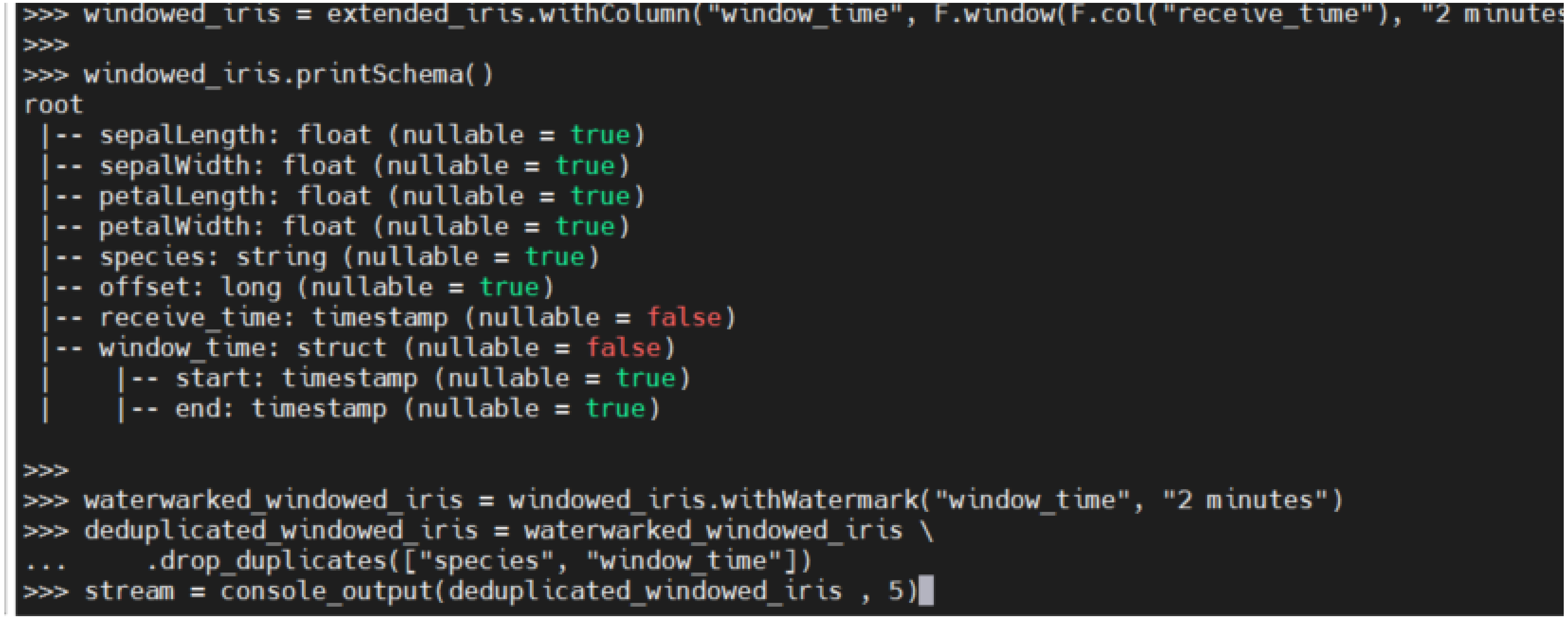
```
[student559_10@bigdataanalytics-worker-3 bin]$ hdfs dfs -ls
Found 5 items
drwx------   - student559_10 student559_10          0 2022-02-01 06:00 .Trash
drwxr-xr-x   - student559_10 student559_10          0 2022-01-28 23:15 .sparkStaging
drwxr-xr-x   - student559_10 student559_10          0 2022-02-05 10:53 checkpoints
drwxr-xr-x   - student559_10 student559_10          0 2022-01-28 14:11 for_stream
drwxr-xr-x   - student559_10 student559_10          0 2022-01-31 10:04 shadrin_iris_kafka_checkpoint
[student559_10@bigdataanalytics-worker-3 bin]$ hdfs dfs -du -h checkpoints/duplicates_console_chk
87     174    checkpoints/duplicates_console_chk/commits
45     90     checkpoints/duplicates_console_chk/metadata
1.3 K  2.6 K  checkpoints/duplicates_console_chk/offsets
41     82     checkpoints/duplicates_console_chk/sources
[student559_10@bigdataanalytics-worker-3 bin]$ 
```



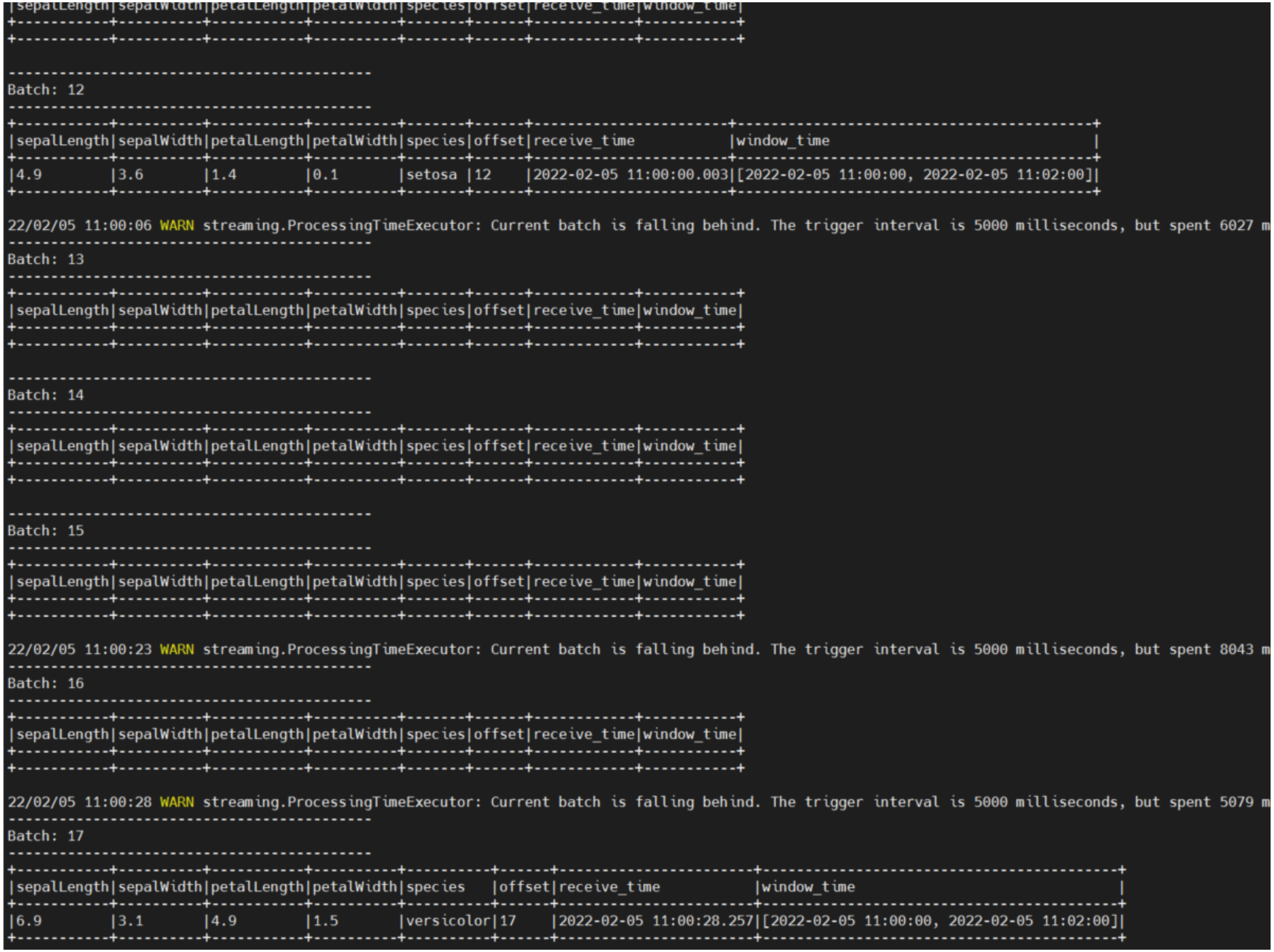
```
>>> deduplicated_iris = waterwarked_iris.drop_duplicates(["species", "receive_time"])
>>> stream = console_output(deduplicated_iris , 20)
-------------------------------------------
Batch: 0
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|4.9        |3.0       |1.4        |0.2       |setosa |0     |2022-02-05 10:55:07.272|
|null       |null      |null       |null      |null   |0     |2022-02-05 10:55:07.272|
+-----------+----------+-----------+----------+-------+------+-----------------------+

-------------------------------------------
Batch: 1
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|5.0        |3.6       |1.4        |0.2       |setosa |1     |2022-02-05 10:55:20.005|
+-----------+----------+-----------+----------+-------+------+-----------------------+

-------------------------------------------
Batch: 2
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|5.0        |3.4       |1.5        |0.2       |setosa |2     |2022-02-05 10:55:40.004|
+-----------+----------+-----------+----------+-------+------+-----------------------+

-------------------------------------------
Batch: 3
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|5.4        |3.7       |1.5        |0.2       |setosa |3     |2022-02-05 10:56:00.005|
+-----------+----------+-----------+----------+-------+------+-----------------------+
```



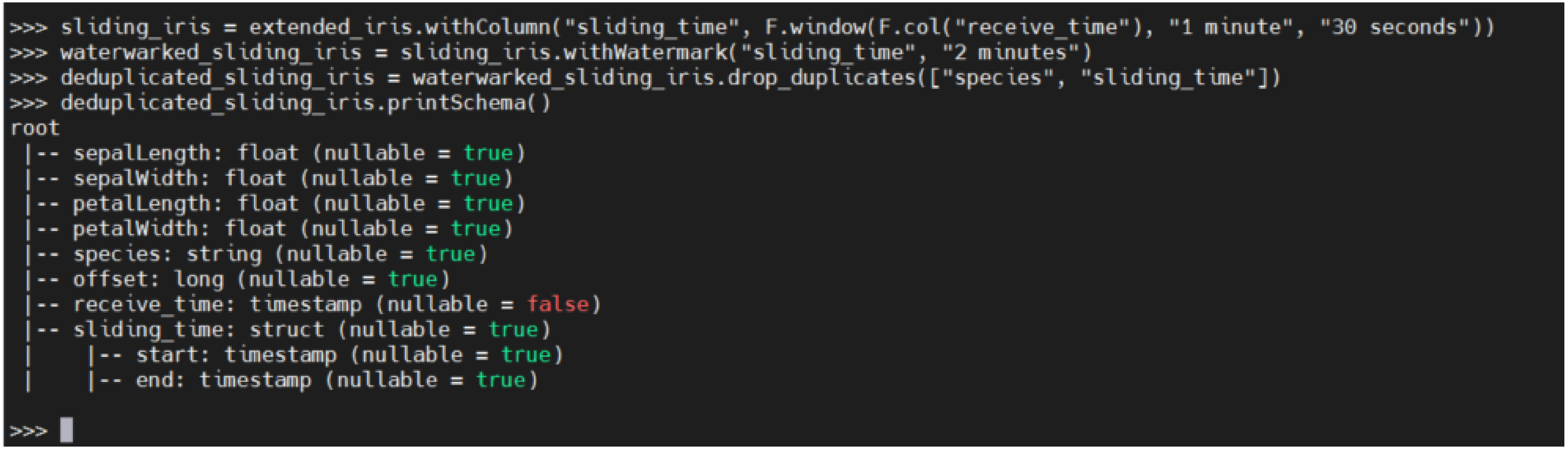
```
 of starting query.
22/02/05 10:57:16 WARN state.HDFSBackedStateStoreProvider: The state for version 4 doesn't exist in
 of starting query.
22/02/05 10:57:16 WARN state.HDFSBackedStateStoreProvider: The state for version 4 doesn't exist in
 of starting query.
[Stage 14:====================================================>  (191 + 1) / 200]22/02/05 10:57:16 W
ot file and delta files if needed...Note that this is normal for the first batch of starting query.
22/02/05 10:57:16 WARN state.HDFSBackedStateStoreProvider: The state for version 4 doesn't exist in
 of starting query.
22/02/05 10:57:16 WARN state.HDFSBackedStateStoreProvider: The state for version 4 doesn't exist in
 of starting query.
22/02/05 10:57:16 WARN state.HDFSBackedStateStoreProvider: The state for version 4 doesn't exist in
 of starting query.
22/02/05 10:57:16 WARN state.HDFSBackedStateStoreProvider: The state for version 4 doesn't exist in
 of starting query.
22/02/05 10:57:16 WARN state.HDFSBackedStateStoreProvider: The state for version 4 doesn't exist in
 of starting query.
22/02/05 10:57:16 WARN state.HDFSBackedStateStoreProvider: The state for version 4 doesn't exist in
 of starting query.
[Stage 14:=====================================================>(198 + 1) / 200]22/02/05 10:57:16 W
ot file and delta files if needed...Note that this is normal for the first batch of starting query.
-------------------------------------------
Batch: 4
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|4.8        |3.4       |1.6        |0.2       |setosa |4     |2022-02-05 10:56:48.266|
+-----------+----------+-----------+----------+-------+------+-----------------------+

-------------------------------------------
Batch: 5
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|5.4        |3.9       |1.3        |0.4       |setosa |5     |2022-02-05 10:57:19.456|
+-----------+----------+-----------+----------+-------+------+-----------------------+

-------------------------------------------
Batch: 6
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |
+-----------+----------+-----------+----------+-------+------+-----------------------+
|5.1        |3.8       |1.5        |0.3       |setosa |6     |2022-02-05 10:57:23.238|
+-----------+----------+-----------+----------+-------+------+-----------------------+
```



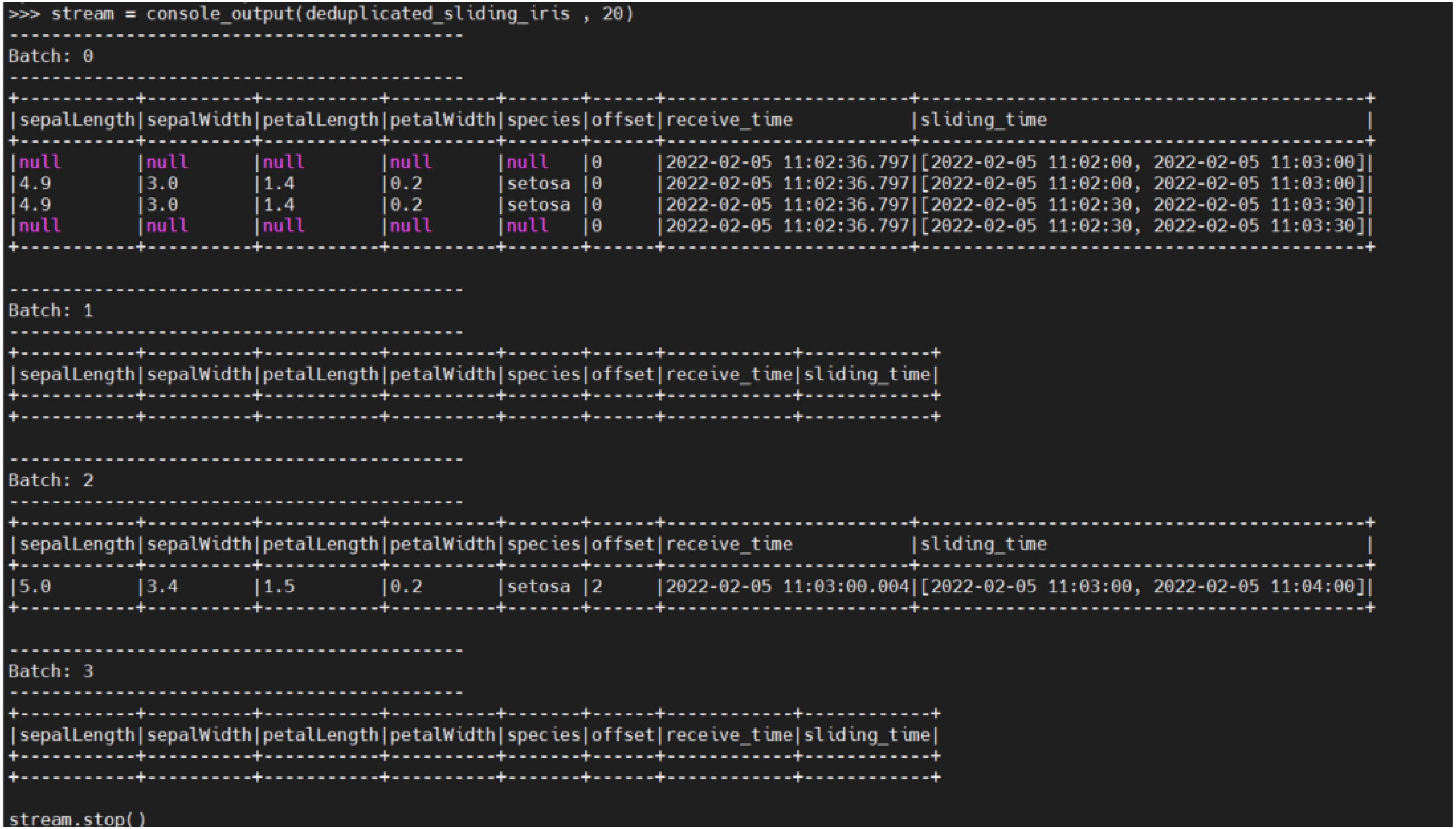
```
>>> windowed_iris = extended_iris.withColumn("window_time", F.window(F.col("receive_time"), "2 minutes
>>>
>>> windowed_iris.printSchema()
root
 |-- sepalLength: float (nullable = true)
 |-- sepalWidth: float (nullable = true)
 |-- petalLength: float (nullable = true)
 |-- petalWidth: float (nullable = true)
 |-- species: string (nullable = true)
 |-- offset: long (nullable = true)
 |-- receive_time: timestamp (nullable = false)
 |-- window_time: struct (nullable = false)
 |    |-- start: timestamp (nullable = true)
 |    |-- end: timestamp (nullable = true)

>>>
>>> waterwarked_windowed_iris = windowed_iris.withWatermark("window_time", "2 minutes")
>>> deduplicated_windowed_iris = waterwarked_windowed_iris \
...     .drop_duplicates(["species", "window_time"])
>>> stream = console_output(deduplicated_windowed_iris , 5)
```



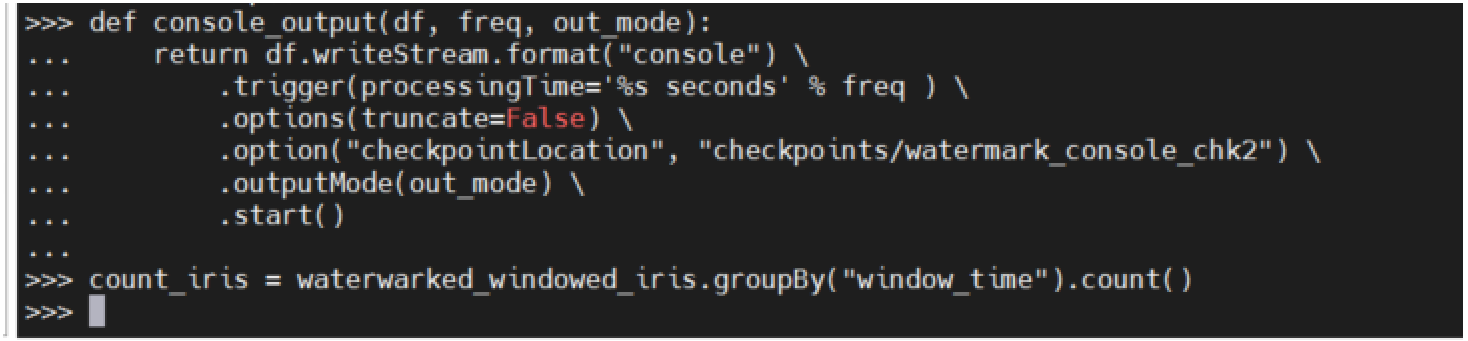
```
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time|window_time|
+-----------+----------+-----------+----------+-------+------+------------+-----------+
+-----------+----------+-----------+----------+-------+------+------------+-----------+

-------------------------------------------
Batch: 12
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+------------------------------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |window_time                               |
+-----------+----------+-----------+----------+-------+------+-----------------------+------------------------------------------+
|4.9        |3.6       |1.4        |0.1       |setosa |12    |2022-02-05 11:00:00.003|[2022-02-05 11:00:00, 2022-02-05 11:02:00]|
+-----------+----------+-----------+----------+-------+------+-----------------------+------------------------------------------+

22/02/05 11:00:06 WARN streaming.ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 5000 milliseconds, but spent 6027 m
-------------------------------------------
Batch: 13
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+------------+-----------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time|window_time|
+-----------+----------+-----------+----------+-------+------+------------+-----------+
+-----------+----------+-----------+----------+-------+------+------------+-----------+

-------------------------------------------
Batch: 14
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+------------+-----------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time|window_time|
+-----------+----------+-----------+----------+-------+------+------------+-----------+
+-----------+----------+-----------+----------+-------+------+------------+-----------+

-------------------------------------------
Batch: 15
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+------------+-----------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time|window_time|
+-----------+----------+-----------+----------+-------+------+------------+-----------+
+-----------+----------+-----------+----------+-------+------+------------+-----------+

22/02/05 11:00:23 WARN streaming.ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 5000 milliseconds, but spent 8043 m
-------------------------------------------
Batch: 16
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+------------+-----------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time|window_time|
+-----------+----------+-----------+----------+-------+------+------------+-----------+
+-----------+----------+-----------+----------+-------+------+------------+-----------+

22/02/05 11:00:28 WARN streaming.ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 5000 milliseconds, but spent 5079 m
-------------------------------------------
Batch: 17
-------------------------------------------
+-----------+----------+-----------+----------+----------+------+-----------------------+------------------------------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species   |offset|receive_time           |window_time                               |
+-----------+----------+-----------+----------+----------+------+-----------------------+------------------------------------------+
|6.9        |3.1       |4.9        |1.5       |versicolor|17    |2022-02-05 11:00:28.257|[2022-02-05 11:00:00, 2022-02-05 11:02:00]|
+-----------+----------+-----------+----------+----------+------+-----------------------+------------------------------------------+
```



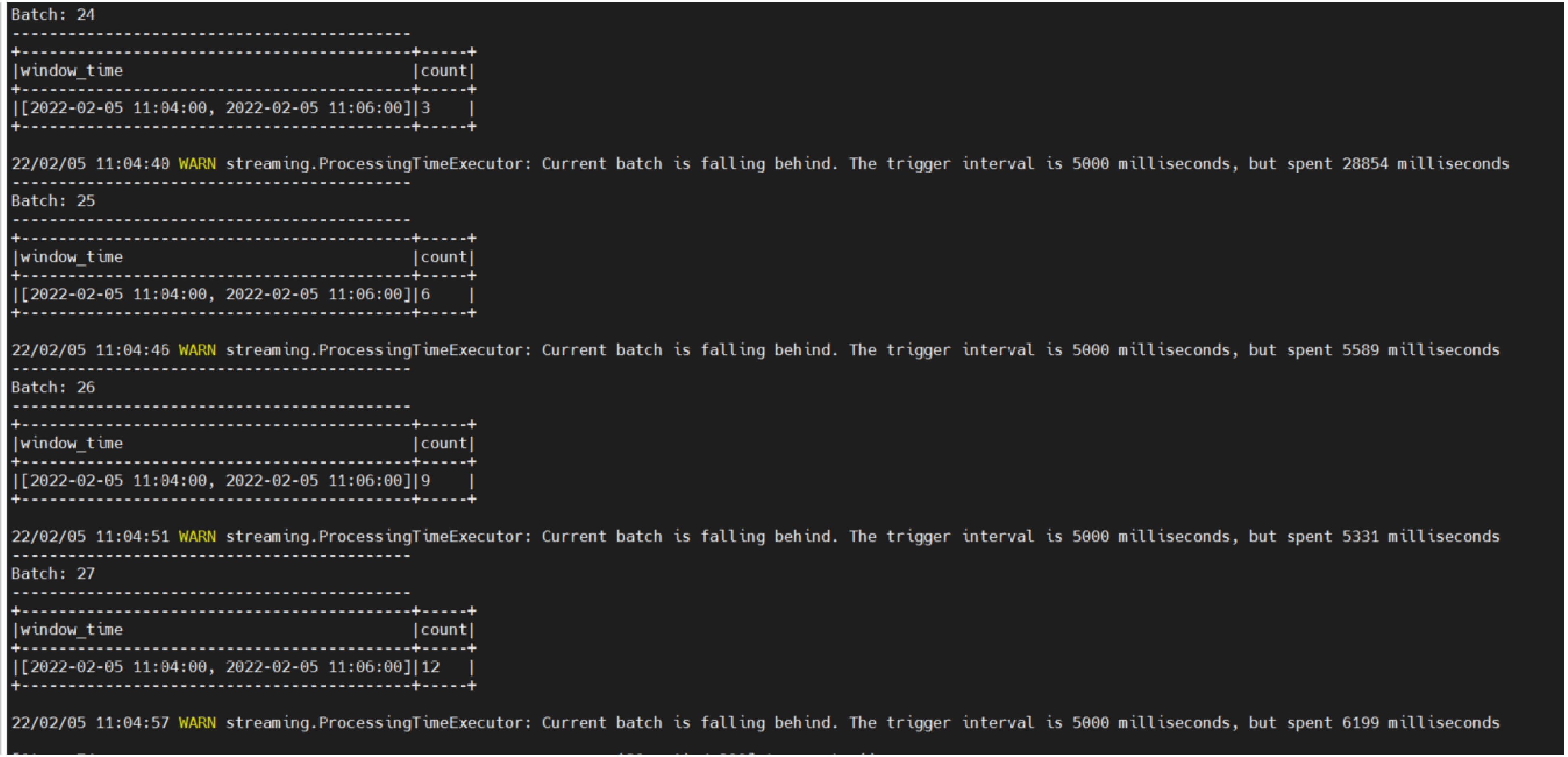
```
>>> sliding_iris = extended_iris.withColumn("sliding_time", F.window(F.col("receive_time"), "1 minute", "30 seconds"))
>>> waterwarked_sliding_iris = sliding_iris.withWatermark("sliding_time", "2 minutes")
>>> deduplicated_sliding_iris = waterwarked_sliding_iris.drop_duplicates(["species", "sliding_time"])
>>> deduplicated_sliding_iris.printSchema()
root
 |-- sepalLength: float (nullable = true)
 |-- sepalWidth: float (nullable = true)
 |-- petalLength: float (nullable = true)
 |-- petalWidth: float (nullable = true)
 |-- species: string (nullable = true)
 |-- offset: long (nullable = true)
 |-- receive_time: timestamp (nullable = false)
 |-- sliding_time: struct (nullable = true)
 |    |-- start: timestamp (nullable = true)
 |    |-- end: timestamp (nullable = true)

>>> 
```



```
>>> stream = console_output(deduplicated_sliding_iris , 20)
-------------------------------------------
Batch: 0
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+------------------------------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |sliding_time                              |
+-----------+----------+-----------+----------+-------+------+-----------------------+------------------------------------------+
|null       |null      |null       |null      |null   |0     |2022-02-05 11:02:36.797|[2022-02-05 11:02:00, 2022-02-05 11:03:00]|
|4.9        |3.0       |1.4        |0.2       |setosa |0     |2022-02-05 11:02:36.797|[2022-02-05 11:02:00, 2022-02-05 11:03:00]|
|4.9        |3.0       |1.4        |0.2       |setosa |0     |2022-02-05 11:02:36.797|[2022-02-05 11:02:30, 2022-02-05 11:03:30]|
|null       |null      |null       |null      |null   |0     |2022-02-05 11:02:36.797|[2022-02-05 11:02:30, 2022-02-05 11:03:30]|
+-----------+----------+-----------+----------+-------+------+-----------------------+------------------------------------------+

-------------------------------------------
Batch: 1
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+------------+------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time|sliding_time|
+-----------+----------+-----------+----------+-------+------+------------+------------+
+-----------+----------+-----------+----------+-------+------+------------+------------+

-------------------------------------------
Batch: 2
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+-----------------------+------------------------------------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time           |sliding_time                              |
+-----------+----------+-----------+----------+-------+------+-----------------------+------------------------------------------+
|5.0        |3.4       |1.5        |0.2       |setosa |2     |2022-02-05 11:03:00.004|[2022-02-05 11:03:00, 2022-02-05 11:04:00]|
+-----------+----------+-----------+----------+-------+------+-----------------------+------------------------------------------+

-------------------------------------------
Batch: 3
-------------------------------------------
+-----------+----------+-----------+----------+-------+------+------------+------------+
|sepalLength|sepalWidth|petalLength|petalWidth|species|offset|receive_time|sliding_time|
+-----------+----------+-----------+----------+-------+------+------------+------------+
+-----------+----------+-----------+----------+-------+------+------------+------------+

stream.stop()
```

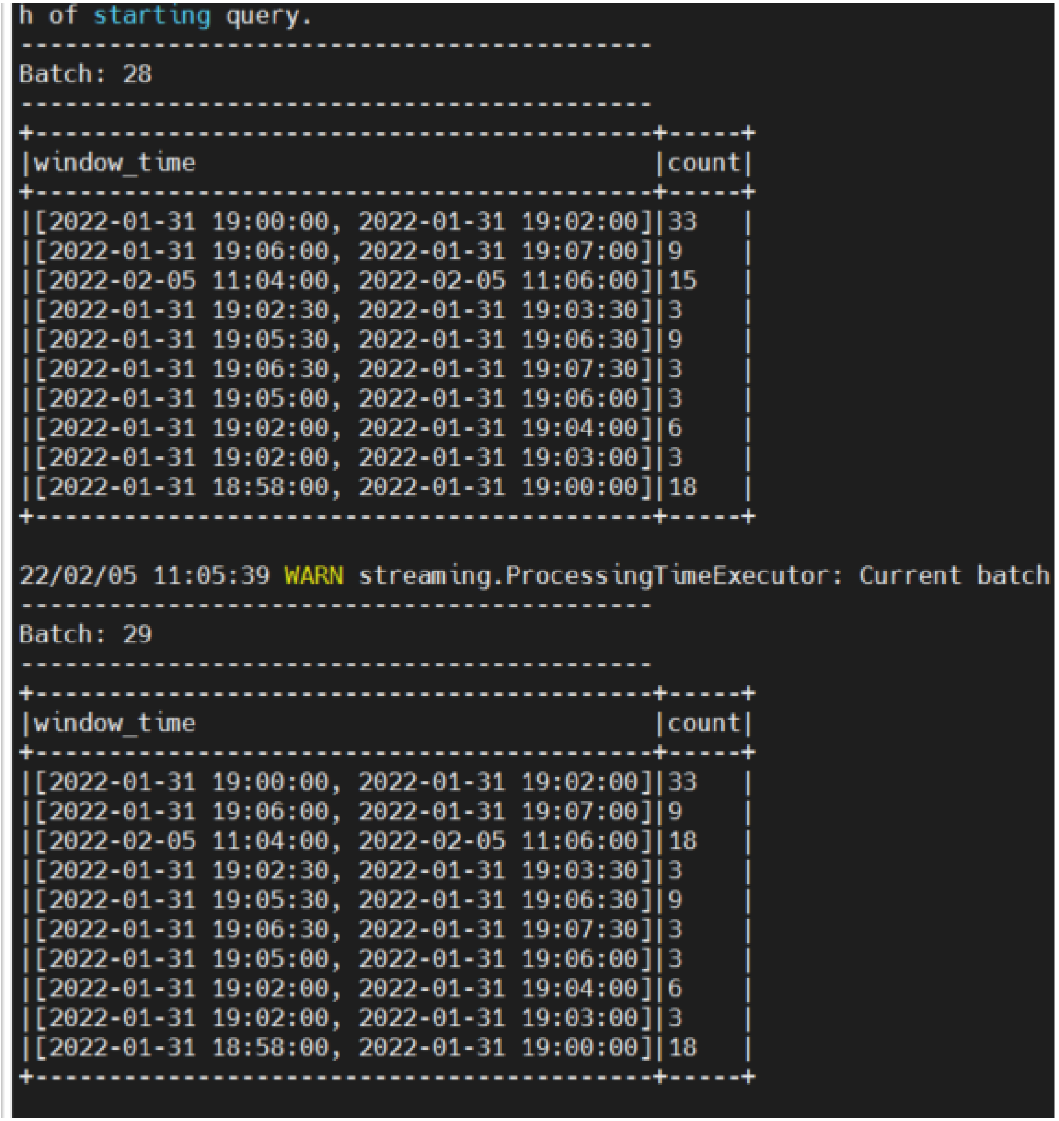


```
>>> def console_output(df, freq, out_mode):
...     return df.writeStream.format("console") \
...         .trigger(processingTime='%s seconds' % freq ) \
...         .options(truncate=False) \
...         .option("checkpointLocation", "checkpoints/watermark_console_chk2") \
...         .outputMode(out_mode) \
...         .start()
...
>>> count_iris = waterwarked_windowed_iris.groupBy("window_time").count()
>>> 
```



```
Batch: 24
-------------------------------------------
+------------------------------------------+-----+
|window_time                               |count|
+------------------------------------------+-----+
|[2022-02-05 11:04:00, 2022-02-05 11:06:00]|3    |
+------------------------------------------+-----+

22/02/05 11:04:40 WARN streaming.ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 5000 milliseconds, but spent 28854 milliseconds
-------------------------------------------
Batch: 25
-------------------------------------------
+------------------------------------------+-----+
|window_time                               |count|
+------------------------------------------+-----+
|[2022-02-05 11:04:00, 2022-02-05 11:06:00]|6    |
+------------------------------------------+-----+

22/02/05 11:04:46 WARN streaming.ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 5000 milliseconds, but spent 5589 milliseconds
-------------------------------------------
Batch: 26
-------------------------------------------
+------------------------------------------+-----+
|window_time                               |count|
+------------------------------------------+-----+
|[2022-02-05 11:04:00, 2022-02-05 11:06:00]|9    |
+------------------------------------------+-----+

22/02/05 11:04:51 WARN streaming.ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 5000 milliseconds, but spent 5331 milliseconds
-------------------------------------------
Batch: 27
-------------------------------------------
+------------------------------------------+-----+
|window_time                               |count|
+------------------------------------------+-----+
|[2022-02-05 11:04:00, 2022-02-05 11:06:00]|12   |
+------------------------------------------+-----+

22/02/05 11:04:57 WARN streaming.ProcessingTimeExecutor: Current batch is falling behind. The trigger interval is 5000 milliseconds, but spent 6199 milliseconds
```

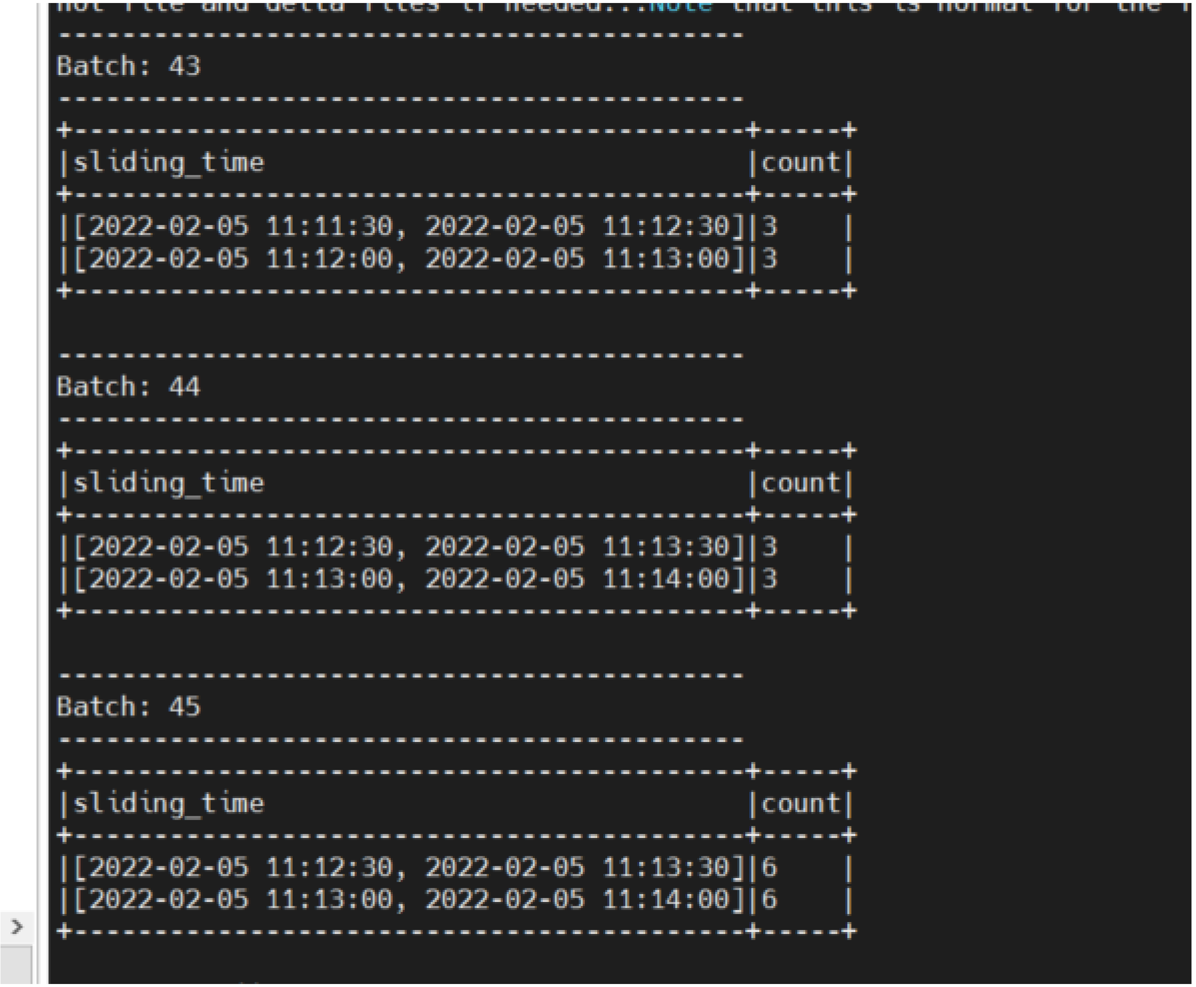
Запустим поток с параметром complete



```
h of starting query.
-------------------------------------------
Batch: 28
-------------------------------------------
+------------------------------------------+-----+
|window_time                               |count|
+------------------------------------------+-----+
|[2022-01-31 19:00:00, 2022-01-31 19:02:00]|33   |
|[2022-01-31 19:06:00, 2022-01-31 19:07:00]|9    |
|[2022-02-05 11:04:00, 2022-02-05 11:06:00]|15   |
|[2022-01-31 19:02:30, 2022-01-31 19:03:30]|3    |
|[2022-01-31 19:05:30, 2022-01-31 19:06:30]|9    |
|[2022-01-31 19:06:30, 2022-01-31 19:07:30]|3    |
|[2022-01-31 19:05:00, 2022-01-31 19:06:00]|3    |
|[2022-01-31 19:02:00, 2022-01-31 19:04:00]|6    |
|[2022-01-31 19:02:00, 2022-01-31 19:03:00]|3    |
|[2022-01-31 18:58:00, 2022-01-31 19:00:00]|18   |
+------------------------------------------+-----+

22/02/05 11:05:39 WARN streaming.ProcessingTimeExecutor: Current batch
-------------------------------------------
Batch: 29
-------------------------------------------
+------------------------------------------+-----+
|window_time                               |count|
+------------------------------------------+-----+
|[2022-01-31 19:00:00, 2022-01-31 19:02:00]|33   |
|[2022-01-31 19:06:00, 2022-01-31 19:07:00]|9    |
|[2022-02-05 11:04:00, 2022-02-05 11:06:00]|18   |
|[2022-01-31 19:02:30, 2022-01-31 19:03:30]|3    |
|[2022-01-31 19:05:30, 2022-01-31 19:06:30]|9    |
|[2022-01-31 19:06:30, 2022-01-31 19:07:30]|3    |
|[2022-01-31 19:05:00, 2022-01-31 19:06:00]|3    |
|[2022-01-31 19:02:00, 2022-01-31 19:04:00]|6    |
|[2022-01-31 19:02:00, 2022-01-31 19:03:00]|3    |
|[2022-01-31 18:58:00, 2022-01-31 19:00:00]|18   |
+------------------------------------------+-----+
```

Запустим поток с параметром append (не работает с агрегирующей функцией count) ➔ пропустим

Запустим sliding



```
not file and delta files if needed...Note that this is normal for the
-------------------------------------------
Batch: 43
-------------------------------------------
+------------------------------------------+-----+
|sliding_time                              |count|
+------------------------------------------+-----+
|[2022-02-05 11:11:30, 2022-02-05 11:12:30]|3    |
|[2022-02-05 11:12:00, 2022-02-05 11:13:00]|3    |
+------------------------------------------+-----+

-------------------------------------------
Batch: 44
-------------------------------------------
+------------------------------------------+-----+
|sliding_time                              |count|
+------------------------------------------+-----+
|[2022-02-05 11:12:30, 2022-02-05 11:13:30]|3    |
|[2022-02-05 11:13:00, 2022-02-05 11:14:00]|3    |
+------------------------------------------+-----+

-------------------------------------------
Batch: 45
-------------------------------------------
+------------------------------------------+-----+
|sliding_time                              |count|
+------------------------------------------+-----+
|[2022-02-05 11:12:30, 2022-02-05 11:13:30]|6    |
|[2022-02-05 11:13:00, 2022-02-05 11:14:00]|6    |
+------------------------------------------+-----+
```